Statistics

Physics 212 C
Dennis Silverman


■ Confidence Intervals

■ Confidence Intervals for Data

\( \text{In[1]} := \text{\textless \textgreater Statistics\'ConfidenceIntervals}' \)

\( \text{In[3]} := \text{data} = \{2.1, 1.2, 0.7, 1.0, 1.1, 3.2, 3.2, 3.3, 2.1, 0.3\}; \)

\( \text{In[4]} := \text{mean} = \text{Mean[data]} \)

\( \text{Out[4]} = 1.82 \)

\( \text{In[7]} := \text{?? VarianceCI} \)

VarianceCI[list] returns a list \{min, max\} representing a confidence interval for the population variance, using the entries in list as a sample drawn from the population.

\text{Attributes[VarianceCI] = \{Protected, ReadProtected\} \)

\text{Options[VarianceCI] = ConfidenceLevel -> 0.95} \)

\( \text{In[8]} := \text{VarianceCI[data]} \)

\( \text{Out[8]} = \{0.593015, 4.1831\} \)

\( \text{In[9]} := \text{VarianceCI[data, ConfidenceLevel -> 0.90]} \)

\( \text{Out[9]} = \{0.667653, 3.39718\} \)

■ Confidence Intervals For Distributions

\( \text{In[10]} := \text{?? NormalCI} \)

NormalCI[mean, sd, ConfidenceLevel -> c] returns a list \{min, max\} representing a confidence interval at confidence level \( c \) for the population mean, based on the sample mean and its standard deviation. This function is used by MeanCI when the population variance is specified.

\text{Attributes[NormalCI] = \{Protected, ReadProtected\} \)

\text{Options[NormalCI] = ConfidenceLevel -> 0.95} \)
\texttt{In[11]}:= \texttt{NormalCI[0, 1]}
\texttt{Out[11]}= \{-1.95996, 1.95996\}

\texttt{In[12]}:= \texttt{NormalCI[0, 1, ConfidenceLevel \rightarrow 0.90]}
\texttt{Out[12]}= \{-1.64485, 1.64485\}

\texttt{In[13]}:= \texttt{?? ChiSquareCI}
\texttt{ChiSquareCI[\text{var, dof, ConfidenceLevel \rightarrow c}] returns a list \{min, max\} representing a confidence interval at confidence level c for the population variance, based on a sample with dof degrees of freedom and unbiased variance estimate \text{var}. This function is used by VarianceCI.}
Attributes[ChiSquareCI] = \{Protected, ReadProtected\}
Options[ChiSquareCI] = ConfidenceLevel \rightarrow 0.95

\texttt{In[14]}:= \texttt{ChiSquareCI[20, 20, ConfidenceLevel \rightarrow 0.67]}
\texttt{Out[14]}= \{15.3701, 28.7551\}

\texttt{In[15]}:= \texttt{ChiSquareCI[20, 20, ConfidenceLevel \rightarrow 0.90]}
\texttt{Out[15]}= \{12.7346, 36.8636\}

\texttt{In[16]}:= \texttt{ChiSquareCI[20, 20]}
\texttt{Out[16]}= \{11.7063, 41.7067\}

\section*{Normal Distribution Functions}

\texttt{In[18]}:= \texttt{<< Statistics`NormalDistribution`}

\texttt{In[19]}:= \texttt{?? ChiSquareDistribution}
\texttt{ChiSquareDistribution[n] represents the chi-square distribution with n degrees of freedom.}

\texttt{In[20]}:= \texttt{?? PDF}
\texttt{PDF[\text{distribution, x}] gives the probability density function of the specified statistical distribution evaluated at \text{x}.}

\texttt{In[21]}:= \texttt{chis = PDF[ChiSquareDistribution[5], x]}
\texttt{Out[21]}= \frac{e^{-x/2} x^{3/2}}{3 \sqrt{2 \pi}}
In[22]:= Plot[chis, {x, 0, 10}]

Out[22]= -Graphics-

In[23]:= ?NormalDistribution

NormalDistribution[μ, σ] represents the normal (Gaussian) distribution with mean 
μ and standard deviation σ.

In[25]:= ndf = PDF[NormalDistribution[5, 2], x]

Out[25]= \frac{e^{-\frac{1}{2} \frac{|x|^2}{\sigma^2}}}{2 \sqrt{2\pi} \sigma}

In[26]:= Plot[ndf, {x, -1, 11}]

Out[26]= -Graphics-

In[27]:= Mean[NormalDistribution[5, 2]]

Out[27]= 5
In[36]:= Variance[NormalDistribution[5, 2]]
Out[36]= 4

In[28]:= StandardDeviation[NormalDistribution[5, 2]]
Out[28]= 2

In[29]:= ? Quantile

Quantile[list, q] gives the q-th quantile of the entries in list (the fraction of entries in list that are less than this value is q). Quantile[distribution, q] gives the q-th quantile of the specified statistical distribution.

In[30]:= Quantile[NormalDistribution[5, 2], .9]
Out[30]= 7.5631

In[31]:= Quantile[NormalDistribution[5, 2], .1]
Out[31]= 2.4369

In[32]:= ? CDF

CDF[distribution, x] gives the cumulative distribution function of the specified statistical distribution evaluated at x. For continuous distributions, this is defined as the integral of the probability density function from the lowest value in the domain to x. For discrete distributions, this is defined as the sum of the probability density function from the lowest value in the domain to x.

In[33]:= nCDF = CDF[NormalDistribution[5, 2], x]

Out[33]= \( \frac{1}{2} \left( 1 + \text{Erf} \left( \frac{-5 + x}{2 \sqrt{2}} \right) \right) \)

In[35]:= Plot[nCDF, {x, -1, 11}]

Out[35]= -Graphics-
Discrete Distributions

\texttt{In[39]}:= \langle\langle \text{Statistics}'\text{DiscreteDistributions}'\rangle\rangle

\texttt{In[40]}:= \text{\textbf{\texttt{BinomialDistribution}}}

BinomialDistribution[n, p] represents the binomial distribution for \(n\) trials with probability \(p\). A BinomialDistribution[n, p] random variable describes the number of successes occurring in \(n\) trials, where the probability of success in each trial is \(p\).

\texttt{In[46]}:= \text{\textbf{\texttt{binompdf = PDF[BinomialDistribution[4, .5], m]}}} \\
\texttt{Out[46]}= 0.0625 \text{ Binomial}[4, m]

\texttt{In[48]}:= \text{\textbf{\texttt{Plot[binompdf, \{m, 0, 6\}]}}} \\
\text{Graphics}

\texttt{In[49]}:= \text{\textbf{\texttt{PoissonDistribution}}}

PoissonDistribution[\(\mu\)] represents the Poisson distribution with mean \(\mu\).

\texttt{In[54]}:= \text{\textbf{\texttt{poissonpdf = PDF[PoissonDistribution[4, m], m]}}} \\
\texttt{Out[54]}= \frac{4^m}{\text{E}^m m!}
Nonlinear Function Fitting

NonlinearFit[data, model, vars, params, (opts)] searches for a least-squares fit to a list of data according to the model containing the variables vars and the parameters params. Parameters may be expressed as a list of symbols or a list of lists, where each parameter is listed with starting value(s) and bounds in one of several different ways: {symbol, start} or {symbol, min, max} or {symbol, {start1, start2}} or {symbol, {start1, start2}, min, max}. The data can have the form {{{x1, y1, ..., f1}, {x2, y2, ..., f2}, ...}, where the number of coordinates x, y, ... is equal to the number of variables in the list vars. The data can also be of the form {f1, f2, ...}, with a single coordinate assumed to take values 1, 2, .... Method specifies the Levenberg-Marquardt or FindMinimum search methods. AccuracyGoal and PrecisionGoal specify the number of digits of absolute and relative error allowed in the residual sum of squares.

data = Table[{n, 3 Cos[2 n]}, {n, 0., 1.5, .5}]

NonlinearFit[data, b Cos[a x], x, {a, b}]

Uses Levenberg-Marquardt Method going from steepest descents to quadratic minimization.